

## Step-by-step Example of Network Analysis Protocol Using Cytoscape

### Setup

This protocol relies on features in Cytoscape<sup>1,2</sup> 3.1.1 and greater, which can be downloaded from <http://www.cytoscape.org>. For analysis of larger AP-MS datasets Cytoscape will require a significant amount of memory and the 64-bit versions of Java and Cytoscape are strongly recommended. 4GB of RAM or more is recommended. All screenshots and steps executed below were obtained on an Apple Macbook Pro with a 2.6GHz Intel Core i7 and 16GB of RAM running Mac OS X version 10.9.2. The use of an SSD drive is also recommended.

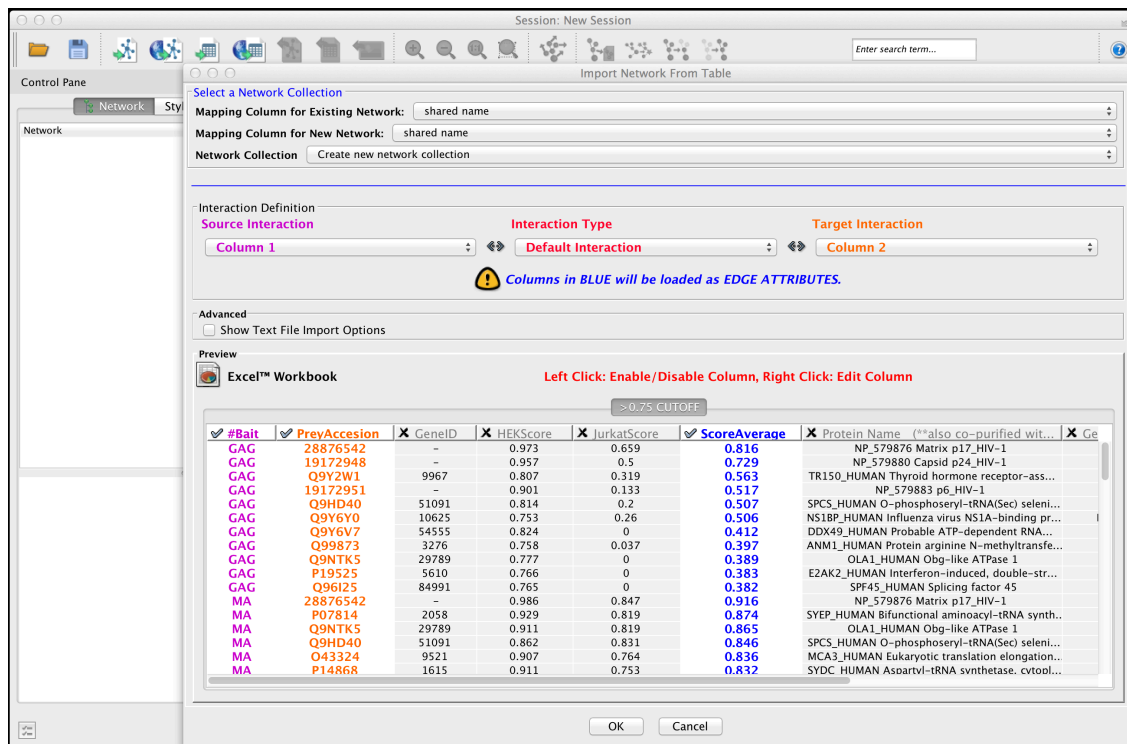
We will also use four apps: BridgeDB, enhancedGraphics, clusterMaker2<sup>3</sup>, and BiNGO<sup>4</sup>. These will be loaded as part of the protocol below.

### Notes

- The procedure below is not meant to be a general tutorial for using Cytoscape or any of the mentioned Apps. A collection of Cytoscape tutorials is available at <http://tutorials.cytoscape.org> for interested readers.
- To help navigate the procedure, all menu items will be specified as a cascade (e.g. **File→Open**), labels will be written in **bold**, and buttons will be *italicized* though they are not italicized in the user interface.
- The low-density figures below all utilize *Supplementary Data 2.xls*, which is taken from Jäger, et al.<sup>5</sup> This data represents the scored data from an AP-MS experiment looking at HIV-human interactions. The Excel file was modified slightly to force numeric prey identifiers to be interpreted as text.
- The high-density figures are taken from Collins, et al.<sup>6</sup> and can be downloaded as a Cytoscape session file as *Supplementary Data 3.cys*.

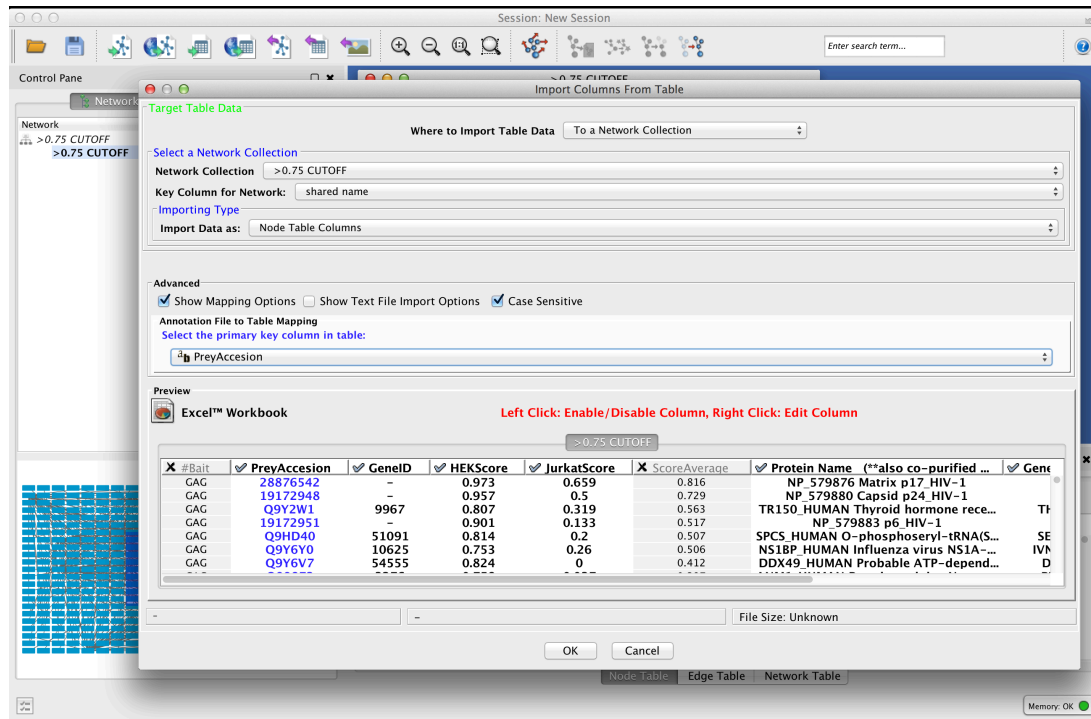
## TUTORIAL STEPS

- 1 Download Cytoscape from <http://www.cytoscape.org/> and install on your computer.
- 2 Importing AP-MS Data
  - a. Import network and interaction (score) data
    - i. Use the network import feature of Cytoscape:  
**File→Import→Network→File...** and select *Supplementary Data 2.xls*, which may be downloaded as part of the main manuscript.
    - ii. In the **Interaction Definition** section, select the bait column (Column 1, #Bait) as your source and the prey column (Column 2, PreyAccession) as target. Any edge attributes (e.g. ScoreAverage) can be imported at this stage by clicking on the column header. Do not import HEKScore and JurkatScore at this stage. We're going to import those as node attributes later on.



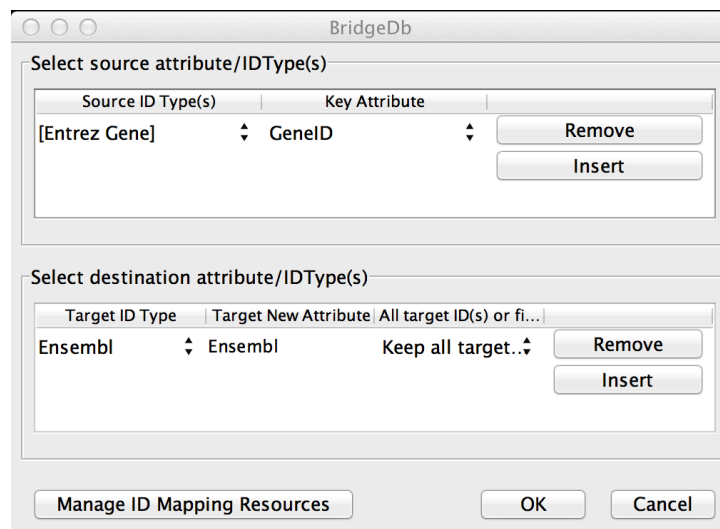
- iii. Click on **OK** to import the network.
- b. Import prey protein information
  - i. Use the table import feature of Cytoscape to import node data:  
**File→Import→Table→File...** and again select *Supplementary Data 2.xls*.
  - ii. Select **Show Mapping Options** and select the prey column (PreyAccession) as the key attribute. By default, all columns are imported. Do not import the bait and any bait protein information or any

of the columns imported in step 2a above. To prevent a column from being imported, click once on its header. The column will not be imported if an “X” mark appears to the left of its name. Click the column headers to disable columns (#Bait and ScoreAverage). For our purposes, we do want to read in the HEKScore and JurkatScore as Node attributes at this point.



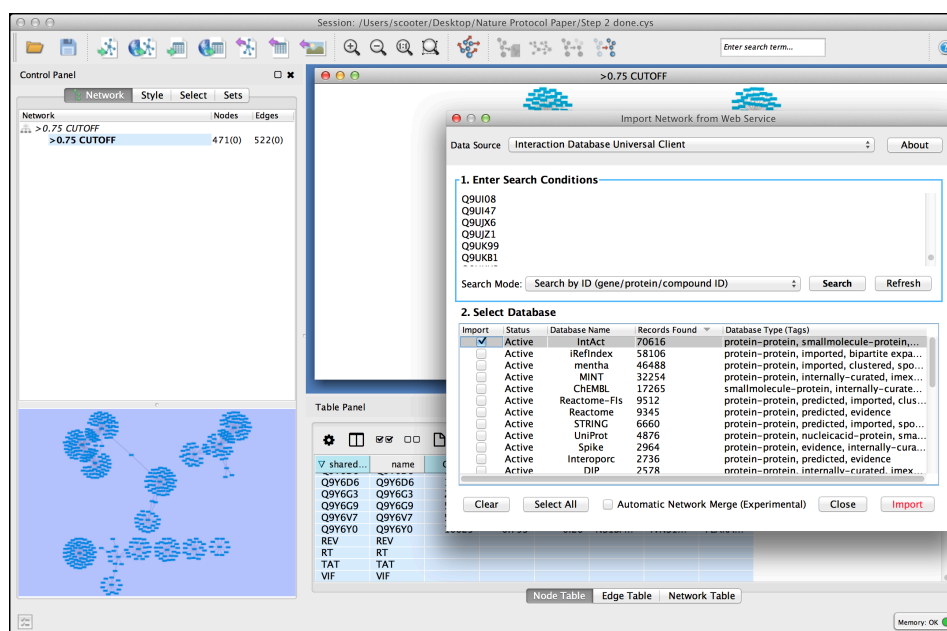
- iii. Click on **OK** to import the attributes.
- c. Import bait protein information (skip this step for our sample since there is no additional information associated only with the bait proteins)
  - i. If we were going to import additional attributes for the bait proteins, we would use the table import feature of Cytoscape to import node data: **File→Import→Table→File...** and for the final time, choose the same data file (*Supplementary Data 2.xls*).
  - ii. Select **Show Mapping** Options and select the bait column (#Bait) as the key attribute. Do not import any of the columns imported in steps 2a and 2b above.
  - iii. Click on **OK** to import the attributes.
- 3 Mapping identifiers (skip this step for our sample since it already includes the identifiers we're going to use).
  - a. If not already loaded, install the bridgeDB App
    - i. Go to **Apps→App Manager**

- ii. In the **Search** box, enter “bridgedb”. Select the app and click on *Install* to install and initialize the BridgeDb App.
  - b. Configure BridgeDb
    - i. **Apps→BridgeDb→Manage ID Mapping Resources**
    - ii. Select *Web Services*
      1. Click once on *Web Services*
      2. Under *Web Service Type*, select BridgeDb web service.
      3. Choose <http://webservice.bridgedb.org/Human> under **Base URL of BridgeDb web service**
      4. Select *OK* to activate the service
  - c. Map from Entrez Gene to Ensembl
    - i. **Apps→BridgeDB→Map Identifiers**
    - ii. GeneID column is populated with Entrez Gene ID’s for all of the prey proteins, so select *Entrez Gene* under **Source ID Type(s)** and select *GeneID* under **Key Attribute**.
    - iii. Create a destination column by selecting *Ensembl* under **Target ID Type** and **Target New Attribute**. Under **All target ID(s) or first only?**, choose **Keep the first target ID only**.
    - iv. Click *OK* to do the mapping

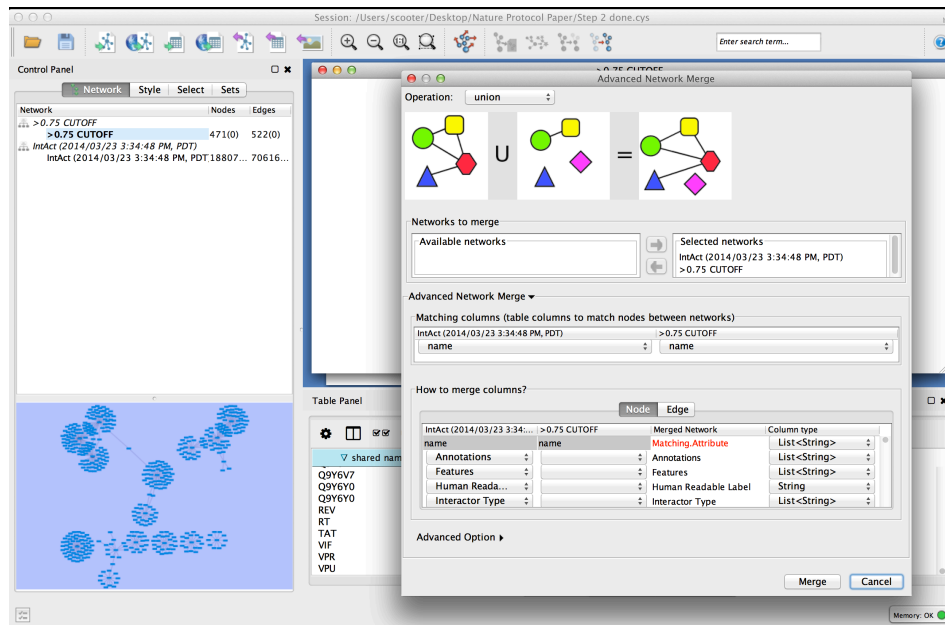


- v. After the mapping is complete, a dialog with **Would you like to map more identifiers?** will appear. Click *No*.
- 4 Enriching data with existing protein-protein interaction data
  - a. Loading public data
    - i. In the **Node Table** panel at the bottom of the Cytoscape desktop, sort the node table for your network by the **name** column by clicking on its header once.

- ii. Expand the **name** column in the table to visually check for cells with multiple identifiers. Edit the cells that include multiple values (e.g. P01892; P04439; etc.) and just choose the first value (you'll find 3 instances of this in our sample). Leave the original specification in **shared name**.
- iii. Bring up the interface to load networks from web services:  
**File→Import→Network→Public Databases....**
- iv. Under **Data Source**, select **Interaction Database Universal Client**
- v. Go back to the **Node Table** panel, and highlight all of the rows in the **name** column of your Node Table by drag or shift-click (note that all of the columns will highlight as well)
- vi. Holding your mouse over a cell in the **name** column, copy the cells using ctrl-C or ⌘-C.
- vii. Paste the resulting names into the **Enter Search Conditions** text box.
  1. Edit the list to remove any entries that are not appropriate identifiers. In our sample case, this would include the names of the HIV proteins (NEF, VIF, GAG, etc.) and the GI numbers of the resulting transcripts (19172951, 25121906, etc.), leaving only UniProt identifiers (O00116, P17066, Q12948, etc).
  2. Click on *Search*.
- viii. Select the database(s) you want to use by clicking on the checkbox under the **Import** column in the **2. Select Database** section. In our case, we're only going to use IntAct, so deselect all of the other databases and make sure only IntAct is selected.



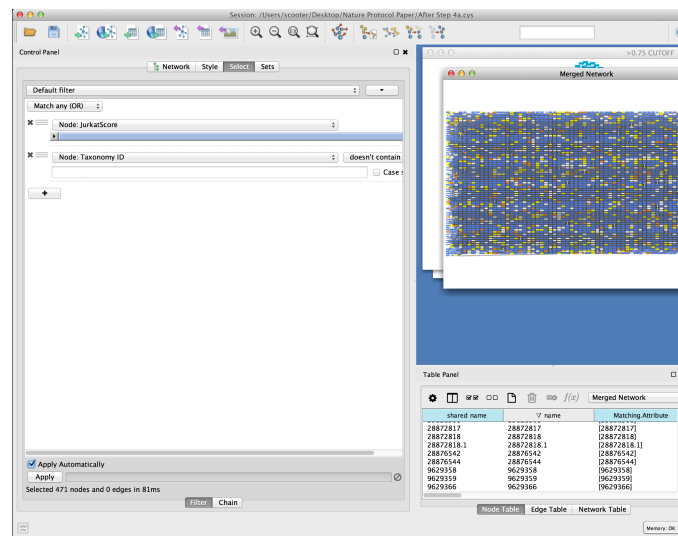
- ix. Click on *Import*. This can take a long time depending on your network bandwidth; 2 minutes on a fast Ethernet connection.
- x. **Do you want to manually merge networks?** Click **Yes**.
- xi. You can now close the Import dialog and proceed to the merge step.
- b. Merge networks and filter results
  - i. If not already prompted, you can bring up the Advanced Network Merge interface: **Tools→Merge→Networks...**
  - ii. Perform a union merge of the original AP-MS network and the public network that was imported by selecting each network to merge in the **Available networks** panel and click the right arrow so that it appears under the **Selected networks** panel.
  - iii. Click *Merge*. This may take some time to complete; 30 seconds on a



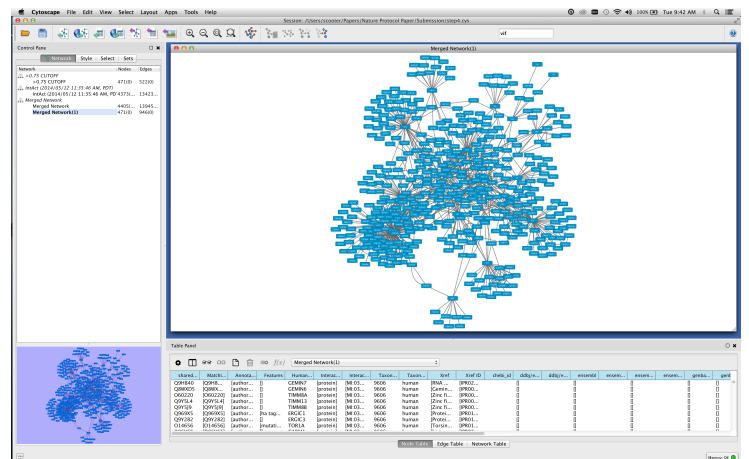
machine with 16GB of RAM.

- iv. The merged network includes a number of additional nodes that we want to remove
- v. Now click the **Select** tab in the **Control Panel**. This will bring up the filters interface that can be used to select the nodes we want.
  1. Because the network is large, uncheck **Apply Automatically** at the bottom of the **Select** panel.
- vi. Click on + then *Column Filter* to add our first filter
  1. Choose *Node: JurkatScore* and then drag the sliders (or enter values) to include all numeric values (e.g., 0 to 1). This will include all of our nodes with a score for Jurkat, and exclude those that are blank.
- vii. Click on + again, then *Column Filter* to add our second filter

1. Choose *Node: Taxonomy ID*. Scroll to the very right of the pull down for the column to find another pull-down for how to do the match (you may need to scroll to the right or expand the filters tab). Select *doesn't contain* and leave the text box blank. This will search for all proteins that do not have anything in the Taxonomy ID field. Most of the nodes in the IntAct network have value for this field, but our HIV proteins do not.
- viii. Change the Boolean type at the top from Match all (AND) to Match any (OR).



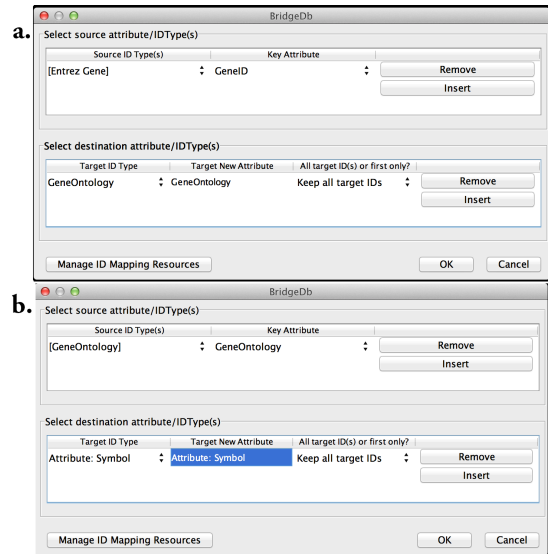
- ix. Now click **Apply**. All of the 471 nodes from the original network should now be selected.
- x. Create a new network **File→New→Network→From selected nodes, all edges**. This will create a new network that all of our original nodes and edges plus the edges that we added from the IntAct database.
- c. If desired, simplify the resulting network
  - i. Remove self-loops (homodimer formation, etc.)
    1. **Edit→Remove Self-Loops**. In the dialog, choose **Merged Network(1)**.
  - ii. Remove duplicated edges (multiple experiments, types of connections)
    1. **Edit→Remove Duplicated Edges**. In the dialog, choose the



same network as above.

## 5 Functional annotation

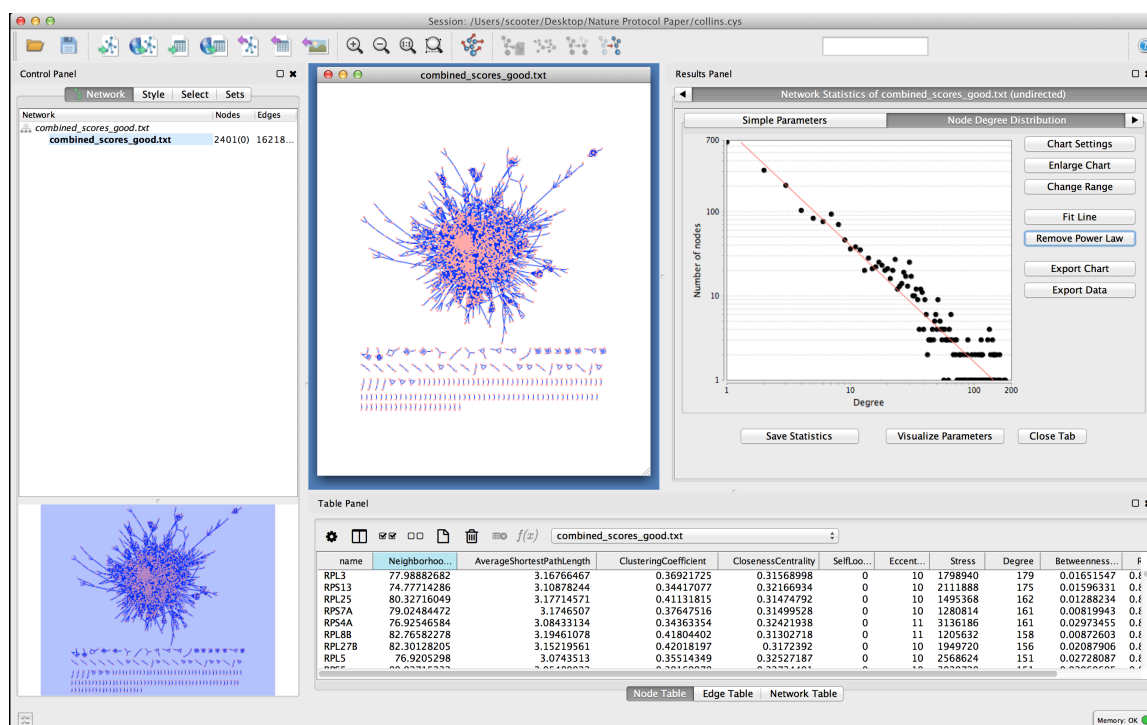
- a. Map GO identifiers to prey proteins using BridgeDb
  - i. **Apps→BridgeDb→Map Identifiers**
  - ii. In the **Source ID Type(s)** pull down, choose **Entrez Gene**. In **Key Attribute**, choose the GeneID column as before.
  - iii. In the **Target ID Type** pull down, choose GeneOntology
  - iv. Click **OK** to do the mapping. With our filtered network of ~450 nodes it only takes 10 seconds on a machine with 16GB of RAM.
  - v. Click **Yes** when prompted to map more identifiers.
- b. Map GO term names to GO identifiers
  - i. Choose **Source ID Type(s)**: GeneOntology, and **Key Attribute**: GeneOntology (i.e., the column created in step 5a)
  - ii. Choose **Target ID Type**: Attribute:Symbol and enter "GeneOntologySymbols" for **Target New Attribute**.
  - iii. Click **OK** to do the mapping. This mapping retrieves GO term names for over 2000 GO term IDs. Mapping to attributes (e.g., Attribute:Symbol) is a much slower function than to other identifiers. This step takes 30 minutes on a machine with 16GB of RAM.
  - iv. Click **No** when prompted to map more identifiers.
- c. Each prey protein now has Gene Ontology terms. We'll determine if any of those terms are statistically significant in step 8 below.
- d. Save the session **File→Save As...** and name the file "Step 5". We'll use this session again for Step 8.



## 6 Network topological analysis (for high-density data).

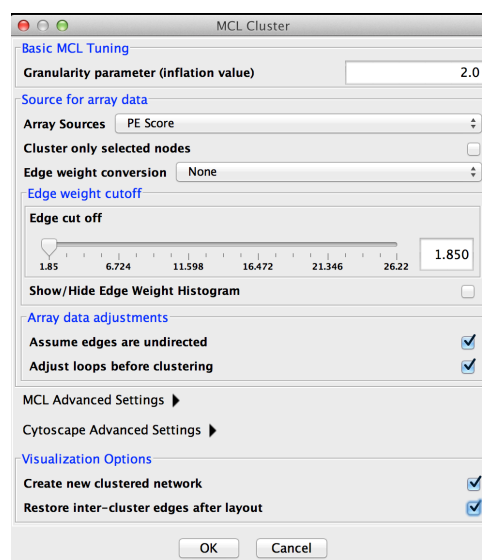
- a. Load *Supplementary Data 3.cys*, which is included in the main manuscript, using **File→Open** and selecting *Supplementary Data 3.cys* from where it was downloaded.
- b. Calculate network statistics using NetworkAnalyzer<sup>7</sup>
  - i. **Tools→NetworkAnalyzer→Network Analysis→Analyze Network**
  - ii. Choose **Treat the network as undirected** and select **OK**
  - iii. A new Results Panel will open with the network statistics. In addition,

new attributes will be created for node degree, closeness centrality, betweenness centrality, stress, and a number of other useful measures.



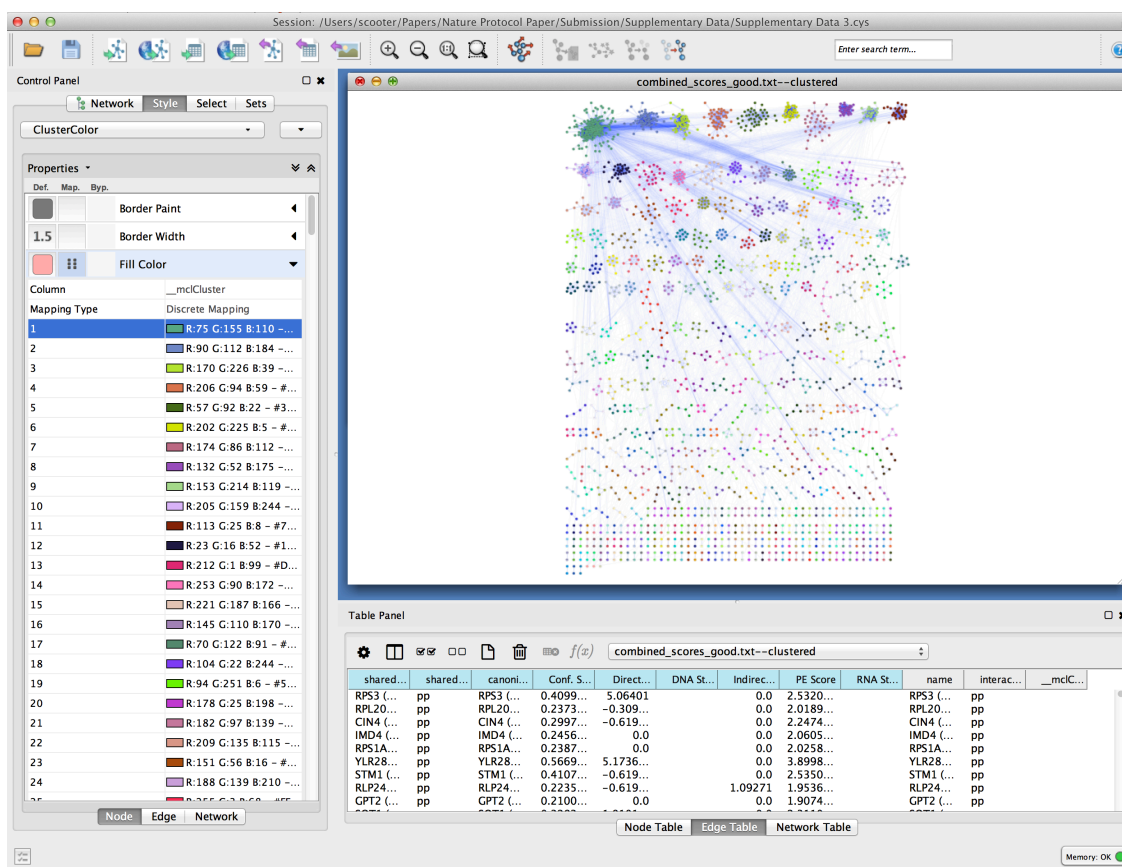
## 7 Clustering (for high-density data)

- If not already loaded, load *Supplementary Data 3.cys* using **File→Open**.
- If not already installed, install clusterMaker2 from the App Manager in the same manner as *bridgeDb App* was loaded in step 3a above.
- MCL Cluster the network to look for complexes
  - Bring up the MCL cluster dialog:  
**Apps→clusterMaker→MCL Cluster**.
  - Set the **Granularity parameter** to 2.0. Generally values between 1.8 and 2.5 are good.
  - Use **PE Score** for the **Array Sources**. This column contains the summary score from the Collins, et al.<sup>6</sup> merger of Krogan<sup>8</sup> and Gavan<sup>9</sup>.
  - In the Visualization Options section, check both **Create New**



## Clustered Network and Restore inter-cluster edges after layout.

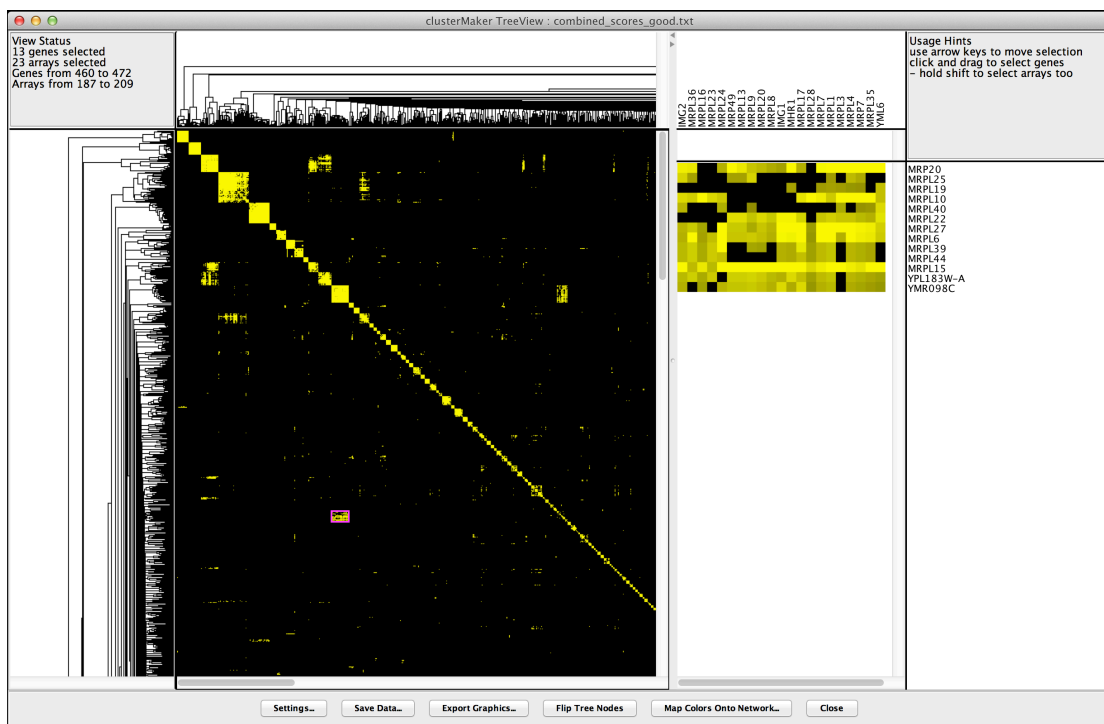
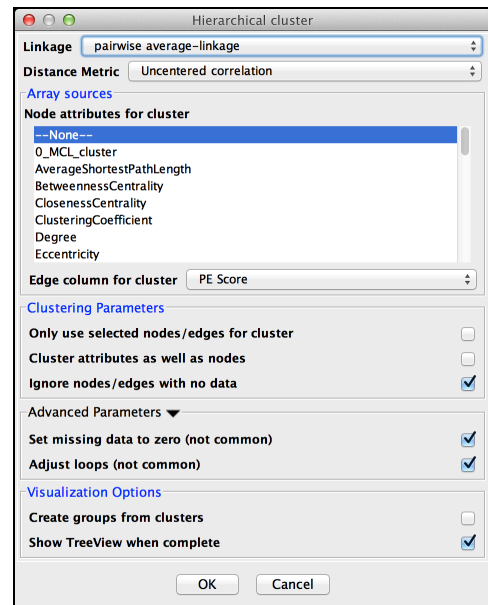
- v. Select **OK** to execute the cluster. You may get an error about the cut-off parameter being out of range. This is normal and may be ignored.
- vi. (Optional) If desired, to highlight the clusters, you can color the nodes according to their cluster.
  1. Go to the **Style** tab.
  2. Open the section labeled **Fill Color** (click on the triangle at the right)
  3. Select **\_\_mclCluster** for the **Column**
  4. Select **Discrete Mapping** for the **Mapping Type**
  5. Right click on any of the vacant cells just created to bring up the context menu and select **Mapping Value Generators**→**Random Color**. Note that your colors might be different than the image



below.

- d. Run a hierarchical cluster to look for complexes and off axis densities.

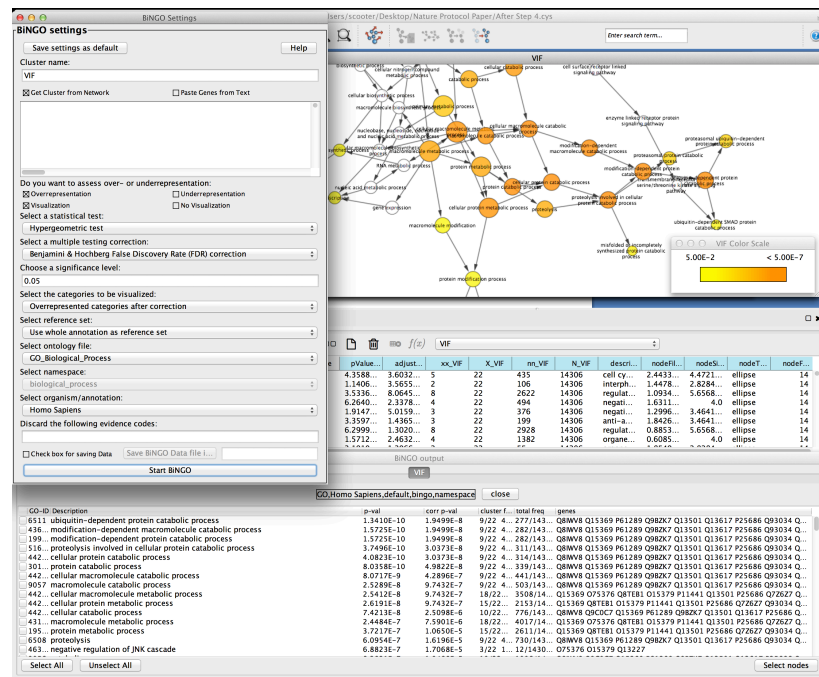
- i. Bring up the Hierarchical cluster dialog:  
**Apps→clusterMaker→Hierarchical Cluster**
- ii. Set the **Distance Metric** to *Uncentered correlation*
- iii. Set the **Edge column for cluster** to *PE Score*
- iv. Under **Advanced Parameters**, click on the checkboxes for **Set missing data to zero** and **Adjust loops**
- v. Click the checkbox for **Show TreeView when complete**
- vi. Press **OK** to create the cluster (shown below)



## 8 Overrepresentation analysis

- a. Install BiNGO<sup>4</sup> from the App store (see step 3a above). Of the two commonly used Cytoscape Apps for GO overrepresentation analysis, BiNGO is a little easier to use but ClueGO<sup>10</sup> provides more options, including the ability to look for enrichment of pathways from KEGG<sup>11,12</sup>, Reactome<sup>13</sup>, and WikiPathways<sup>14,15</sup>. See <http://www.ici.upmc.fr/cluego/>

- for detailed instructions on ClueGO.
- b. Load the session saved at the end of Step 5 above using **File→Open...**
    - i. Click **OK** to remove the current networks and load the new session.
  - c. Select one bait protein (e.g. VIF) and then select all of its prey proteins
    - i. **Select→Nodes→First neighbors of selected nodes→Undirected**, or use accelerator key.
  - d. Activate BiNGO
    - i. **Apps→BiNGO**
    - ii. Enter the name of selected bait protein (e.g., VIF) into BiNGO panel field **Cluster name**
    - iii. Choose Homo Sapiens from **Select organism/annotation**
    - iv. Press **Start BiNGO**
    - v. Observe results to determine the statistically significant GO terms for the protein (see the figure below)



- vi. Selecting an specific GO term in the **BiNGO Output** panel and pressing **Select Nodes** will select those nodes in the original network with that term
- e. Repeat step 6c for each bait protein of interest

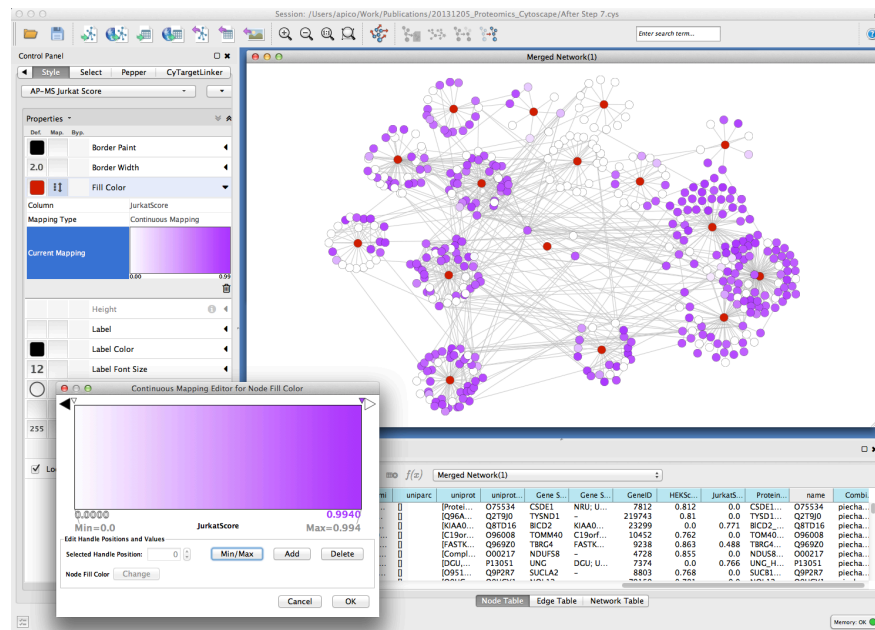
## 9 Visualizing results

- a. Create a Style for Jurkat Scores
  - i. Go to the **Style** tab in the **Control Panel**
  - ii. Via the pull-down list to right of the **Current Style**, choose **Create New**

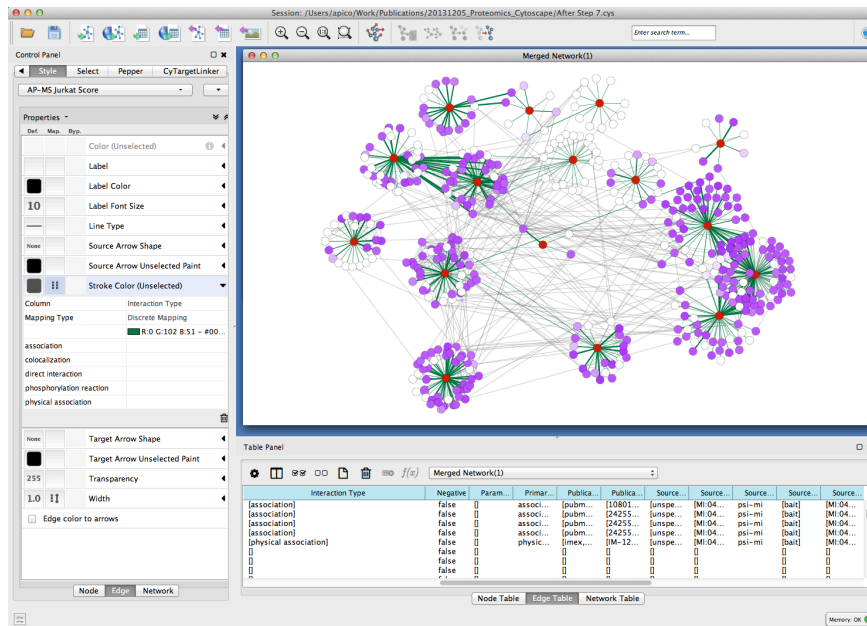
*Visual Style*, name it "AP-MS Jurkat Score"

iii. In the **Properties** table, click on the field for mapping value (**Map.**) for **Fill Color**

1. Choose **Column:** *JurkatScore*
2. Choose **Mapping Type:** *Continuous Mapping*
3. Double click on **Current Mapping** gradient image to open the Editor dialog. Double click on handles along top of gradient to choose colors. Click and drag handles to control gradient.



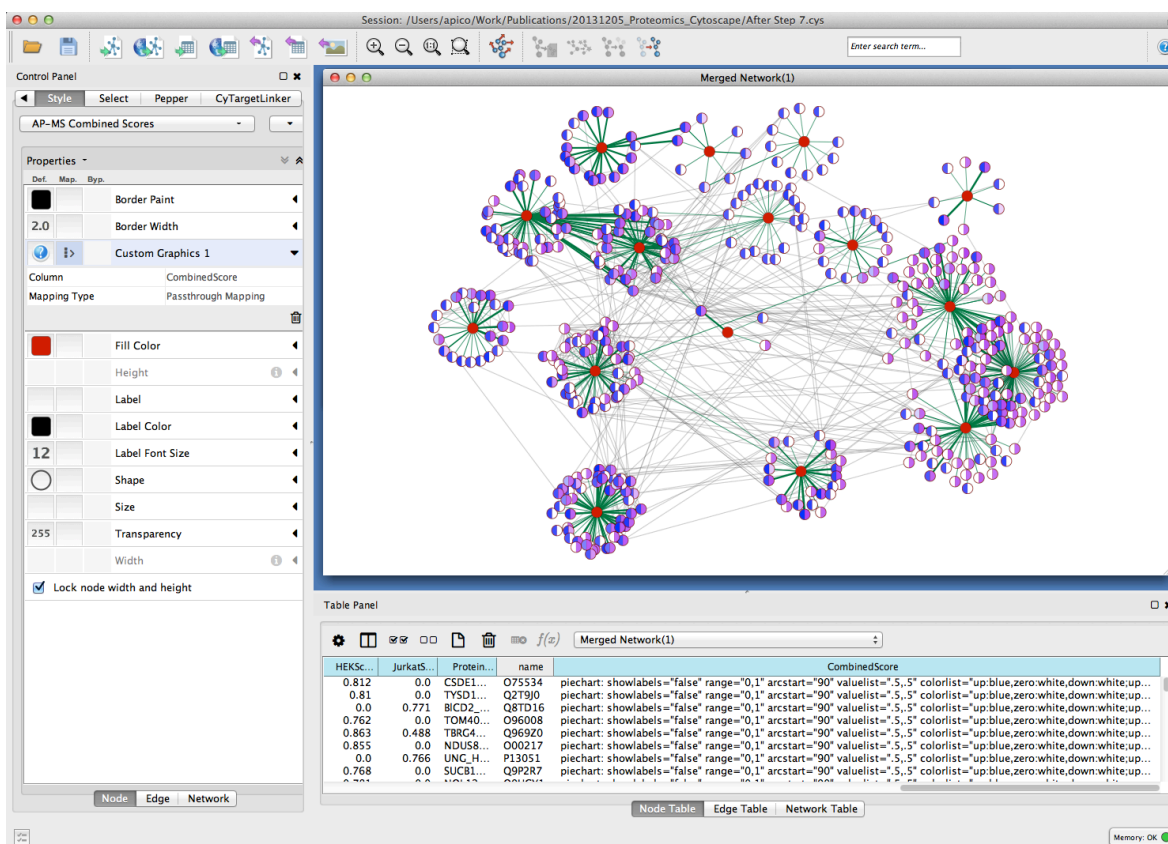
- iv. Select **Edge** tab at bottom of **Style** panel
- v. In the **Properties** table, click on the mapping value (**Map.**) for **Width**
  1. Choose **Column:** *ScoreAverage*
  2. Choose **Mapping Type:** *Continuous Mapping*
  3. Double click on **Current Mapping** gradient image to open the Editor dialog. Click and drag handles to control gradient
- vi. In the **Properties** table, click on the mapping value (**Map.**) for **Stroke Color**
  1. Choose **Column:** *Interaction Type*
  2. Choose **Mapping Type:** *Discrete Mapping*
  3. Double click in right column next to values to choose colors. Leave empty to use default color. Note: these values are from the public data we imported in step 4, so the blank row represents our original data, for which you can nevertheless set a color.



- b. (Optional) Create a second visual style for HEK Scores from a copy
  - i. Via the pull-down list to right of the **Current Style**, choose *Copy Visual Style*, name it "AP-MS HEK Score"
  - ii. View the expanded settings for **Fill Color**
    1. Choose **Column: HEKScore**
    2. Double click gradient image to select alternate colors
- c. Create a combined visual style with enhancedGraphics
  - i. First, load the enhancedGraphics app from the **App Store** (see step 3a above).
  - ii. In the **Table Panel**, click the button for *Create New Column* (looks like a blank page), , then select New Single Column→String to create a new String column.
  - iii. Enter "CombinedScore" in the *Please enter new column name* field.
  - iv. Select the first cell in the newly created column and paste the following text:
 

```
piechart: showlabels="false" range="0,1" arcstart="90" valuelist=".5,.5"
colorlist="up:blue,zero:white,down:white;up:purple,zero:white,
down:white" attributelist="HEKScore,JurkatScore"
labellist="HEKScore,JurkatScore"
```
  - v. Right click on cell and select *Apply to entire column*
  - vi. Return to the **Style** panel and choose *Copy Style* again from the pull-down to the right and name it "AP-MS Combined Scores"
  - vii. In the **Properties** table, select the pull-down menu and choose *Show All*
  - viii. Click on the mapping value (**Map.**) for **Custom Graphics 1**

- ix. Choose **Column: CombinedScore**
- x. Choose **Mapping Type: Passthrough Mapping**



- d. Exporting results as graphics
  - i. **File→Export→Network View as Graphics...**
  - ii. Select **PDF File** (or **SVG**) for vector graphics
  - iii. Choose save location and click **OK**

## References

- 1 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
- 2 Cline, M. S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nature protocols* **2**, 2366-2382, doi:10.1038/nprot.2007.324 (2007).
- 3 Morris, J. H. *et al.* clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC bioinformatics* **12**, 436, doi:10.1186/1471-2105-12-436 (2011).
- 4 Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448-3449, doi:10.1093/bioinformatics/bti551 (2005).
- 5 Jager, S. *et al.* Global landscape of HIV-human protein complexes. *Nature* **481**, 365-370, doi:10.1038/nature10719 (2012).
- 6 Collins, S. R. *et al.* Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Molecular & cellular proteomics : MCP* **6**, 439-450, doi:10.1074/mcp.M600381-MCP200 (2007).
- 7 Doncheva, N. T., Assenov, Y., Domingues, F. S. & Albrecht, M. Topological analysis and interactive visualization of biological networks and protein structures. *Nature protocols* **7**, 670-685, doi:10.1038/nprot.2012.004 (2012).
- 8 Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637-643, doi:10.1038/nature04670 (2006).
- 9 Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631-636, doi:10.1038/nature04532 (2006).
- 10 Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093, doi:10.1093/bioinformatics/btp101 (2009).
- 11 Kanehisa, M. *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research* **34**, D354-357, doi:10.1093/nar/gkj102 (2006).
- 12 Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic acids research* **32**, D277-280, doi:10.1093/nar/gkh063 (2004).
- 13 Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic acids research* **42**, D472-477, doi:10.1093/nar/gkt1102 (2014).
- 14 Pico, A. R. *et al.* WikiPathways: pathway editing for the people. *PLoS biology* **6**, e184, doi:10.1371/journal.pbio.0060184 (2008).
- 15 Kelder, T. *et al.* WikiPathways: building research communities on biological pathways. *Nucleic acids research* **40**, D1301-1307, doi:10.1093/nar/gkr1074 (2012).